

Improving Efficiency Through Consolidation of Jurisdictions? Evidence from the Cantons of Switzerland

Philippe K. Widmer^{a,*}, George Elias^b, Peter Zweifel^a

^a*University of Zurich, Department of Economics, Switzerland*

^b*University of Bern, Department of Economics, Switzerland*

Abstract

The purpose of this paper is to analyze the optimal scale of local jurisdictions (cantons) in Switzerland applying Data Envelopment Analysis (DEA) to the years 2000 to 2004. Aggregate output performance indicators for four local government activities (administration, education, health, and transportation) are used to measure technical and scale efficiency and to derive DEA scores. Results show that these public services fail to exhibit economies of scale, undermining quests for centralization of public good provision while suggesting the possibility of Tiebout competition.

Keywords: DEA, efficiency measurement, economies of scale, public good provision, Switzerland,

JEL: C14, C67, H11, H72, H83

* Corresponding author: Philippe Widmer, Department of Economics, University of Zurich, Hottingerstrasse 10, CH-8032 Zurich, Switzerland. Phone: +41 (0)79 732 50 42, Fax: +41 (0)62 205 15 80 87, E-mail: philippe.widmer@polynomics.ch

1. Introduction

The assignment of responsibilities to levels of government is an ongoing issue among federal states worldwide. Some countries have a tradition of transferring additional responsibilities to their member states while others have pursued centralization or even have been delegating authority to supra-national institutions, as in the case of the European Union [see e.g. Anwar (2004) for an overview of recent trends]. These diverging developments give rise to the question, "What allocation of authorities and responsibilities is most likely to maximize national welfare?".

Economists have been responding to this question with prescriptions regarding the optimal allocation of public resources [see e.g. Baleiras (2001) and Gómez (2010)]. Two pioneering contributions are those by Olson (1969) and Oates (1972), which state two principles. First, a political jurisdiction must coincide with the geographical area that benefits from its public services (correspondence principle). Second, the marginal benefit of provision must equal its marginal cost in each jurisdiction, precluding both beneficial spillovers and free riding (equivalence principle). This also means that a jurisdiction's authority to tax needs to be congruent with its authority to spend. Oates (1972) adds a third principle, that of decentralization. It calls for small jurisdictions to ensure that public good provision reflects the trade-off between the improved matching of local consumer preferences and the extra cost of differentiation caused by production at a small scale. In the borderline case of constant economies to scale, Oates (2006) still argues in favor of decentralization in the provision of public goods. In the case of positive economies to scale, the associated cost advantage needs to be weighed against the degree of diversity in citizens' preferences.

However, especially in countries where public resources are becoming scarce because of financial distress or increased tax competition, the overriding objective is to increase cost efficiency when considering any change in the division of labor between the federal level and member states. Expectations of increasing economies of scale are used as key arguments in favor of centralization and consolidation of local jurisdictions into larger entities. In Germany for instance, local jurisdictions have been forced to merge in the hope of achieving cost savings. In Switzerland, such mergers have occurred with the promise of a lowered tax burden [for countries outside Europe, see Anwar (2004)]. Evidently, politicians currently focus on economies of scale while neglecting the potential welfare loss due to a smaller consumer surplus caused by

failure to meet citizens' preferences.

Exclusive emphasis on economies of scale is mistaken for an additional reason. It is quite conceivable that one type of public service exhibits economies of scale while another has diseconomies of scale. In this event, the optimal amount of provision depends on the public good considered, precluding the determination of *the* optimal size of a jurisdiction. As soon as the diseconomies of scale have a higher effect on public expenditure than economies of scale, consolidation of jurisdictions would result in higher overall cost of service. The more recent literature on fiscal federalism therefore argues in favor of functional, overlapping, and competing jurisdictions [FOCJ, see Frey (2005)]. In this case, jurisdiction size is not optimized as a whole but individually for each service. If a service indeed exhibits economies of scale, it can either be centralized at the next higher level or be provided in collaboration with jurisdictions whose citizens have similar preferences. In doing so, jurisdictional structure becomes more flexible, reflecting preferences for types of public goods and economies of scale in their provision.

This paper revolves around the choice between consolidation and independent provision of public services by 26 Swiss cantons¹. The objective is to determine the optimal scale for four local government activities (administration, education, health, and transportation). Using annual data covering the years 2000 to 2004, Data Envelopment Analysis (DEA) is applied in order to derive technical and scale efficiency scores for each canton. The crucial issue is whether these four services are characterized by increasing or decreasing economies of scale. In view of the fact that the populations of Swiss cantons continue to differ in their preferences regarding public goods [Kellermann (2007)], it would take strong evidence of increasing returns to scale to justify the quest for additional mergers by federal government.

Results show that in these four public services fail to exhibit economies of scale. Therefore, there is no 'natural monopoly' argument here, generating scope for Tiebout competition between the cantons.

This paper is organized as follows. Section 2 explains Data Envelopment Analysis and the derivation of cost efficiency and scale efficiency scores. Section 3 outlines the political system

¹ The 26 Swiss cantons are Appenzell Inner-Rhodes (AI), Appenzell Outer-Rhodes (AR), Argovia (AG), Basel-City (BS), Basel-Country (BL), Bern (BE), Fribourg (FR), Geneva (GE), Glarus (GL), Grisons (GR), Jura (JU), Lucerne (LU), Neuchatel (NE), Nidwalden (NW), Obwalden (OW), Schaffhausen (SH), Schwyz (SZ), Solothurn (SO), St.Gallen (SG), Thurgovia (TG), Ticino (TI), Uri (UR), Valais (VS), Vaud (VD), Zug (ZG), and Zurich (ZH).

of Switzerland and describes the data used. Section 4 presents efficiency scores of the cantons with regard to the four services as well as tests for nonconstant returns to scale and robustness checks. The paper concludes with policy implications and an outlook to further research in Section 5.

2. Measuring Technical and Scale Efficiency Using Data Envelopment Analysis

Swiss cantons justifiably can be viewed as largely independent producers of a subset of public goods [see also Widmer and Zweifel (2012)]. For productivity measurement, they constitute decision making units (DMU) that transform inputs into outputs, with productivity reflecting the quality of their administration. Following Koopmans (1951), the production process can be measured by a technology set $\Gamma = \{(X, Y) | Y \leq f(X)\}$, describing all possible input and output combinations (X, Y) . However, combinations of interest are only those on the boundary of Γ , represented in (1) by an input and output isoquant $Iso(\cdot)$,

$$\begin{aligned} Iso\ X(y) &= \{x | x \in X(y), \theta x \notin X(y), \forall 0 < \theta < 1\} \\ Iso\ Y(x) &= \{y | y \in Y(x), \theta^{-1}y \notin Y(x), \forall 0 < \theta < 1\}. \end{aligned} \quad (1)$$

The isoquants $Iso(\cdot)$ define sections with strong and weak technical efficiency, depending on the slope of the frontier. They are strongly efficient in the technical sense if it is not possible to produce more outputs Y for a given set of input quantities \bar{X} (output orientation), or conversely, use smaller quantities of inputs X to produce a given set of output quantities \bar{Y} (input orientation). In the input orientation, θ denotes a scalar by which all inputs can be reduced without leaving the feasibility set; accordingly, $\theta = 1$ indicates full technical efficiency TE_I [see Farrell (1957)]. In the output-orientated measure of technical efficiency (TE_o), the scaling-up factor is symbolized by θ^{-1} ,

$$\begin{aligned} TE_I(y, x) &= \min\{\theta : \theta x \in X(y)\}, \text{ for } x \in X(y), TE_I = 1; \\ TE_o(y, x) &= \min\{\theta^{-1} : \theta^{-1}y \in Y(x)\}, \text{ for } y \in Y(x), TE_o = 1. \end{aligned} \quad (2)$$

These efficiency measures are part of the solution to an optimization problem that is usually called Data Envelopment Analysis (DEA) pioneered by Charnes et al. (1979) and subsequently

modified by Banker et al. (1984), among others. The DEA version employed here is an input-orientated one that allows for robust efficiency scores. Robust DEA efficiency scores are measured by first detecting outliers using the concept of superefficiency [see Andersen and Petersen (1993)]. Superefficiency means that a canton is identified as an outlier and excluded from the technology set Γ when its super-efficiency score exceeds the 1.5-fold of the interquartile distance. DEA scores are then taken in a second step as an input to a bootstrapping algorithm developed by Simar and Wilson (2000) in order to derive their asymptotic distribution.

Equation (3) conveys the basic idea of DEA. The objective is to determine an efficient frontier $\widehat{IsoX}(y)$ that is defined by the most productive DMUs. It describes a linear optimization problem for a particular canton c , where Y and X are the output and input matrices and s^+ and s^- , output and input slack parameters with ε as a penalty parameter for the existence of slacks [for more detail, see e.g. Bogetoft and Otto (2011)]. DEA permits to localize the Pareto-efficient efficiency frontier as well as to derive an efficiency score for each canton on the weakly efficient part of $\widehat{IsoX}(y)$,

$$\begin{aligned}
& \text{Min}_{\lambda, s^+, s^-} && \theta_c - \varepsilon \left(\sum_{r=1}^R s_r^+ + \sum_{m=1}^M s_m^- \right) \\
& \text{s.t.} && \sum_{c=1}^C \lambda_c y_{r,c} - s_r^+ = y_{r,c}, \quad r = 1, \dots, R \\
& && \sum_{c=1}^C \lambda_c x_{m,c} - s_m^- = \theta_c x_{m,c}, \quad m = 1, \dots, M \\
& && \sum_c \lambda_c \geq 0 \quad (\text{CRS}); \\
& && \sum_c \lambda_c = 1 \quad (\text{VRS}).
\end{aligned} \tag{3}$$

Here, C is the number of cantons, R the number of outputs, and M the number of inputs. Given the optimization problem, a canton minimizes its inputs X_c for given outputs Y_c by reducing θ_c as long as the scaled inputs $\theta_c X_c$ are equal to $\sum_{c=1}^n \lambda_c x_{i,c} - s_i^-$. The scaling factor θ is a measure of the radial input efficiency of DMU_c in that it reflects the proportion by which all observed inputs can be reduced without reducing the level of outputs. Weights λ_c serve to aggregate both outputs and inputs; note that they do not depend on canton c .

Depending on the restriction imposed on λ_c in equation (3), assumptions regarding economies

of scale can be introduced. Efficiency scores for constant (CRS) and variable (VRS) economies of scale can be calculated using the constraints in (3). The linear optimization problem is solved for each canton in the sample, $c = 1, \dots, C$, applying either the CRS or the VRS constraint.

One of the benefits of this approach is that scale efficiency SE_c can be easily derived from the VRS and CRS scores [see Bogetoft and Otto (2011)]. For canton c , one has,

$$SE_c(y, x) = \frac{\text{CRS efficiency of } DMU_c}{\text{VRS efficiency of } DMU_c} \quad (4)$$

Because the CRS efficiency score can never exceed the VRS value, scale efficiency is bounded by $0 \leq SE_c \leq 1$. Thus, the larger the divergence between VRS and CRS, the lower is the SE_c value. A canton has reached its optimal size if it has a scale efficiency equal to one. Furthermore, the direction of scale adjustment can be determined from the results of the CRS model. Increasing economies of scale exist for cantons with $\lambda_c < 1$, while $\lambda_c > 1$ indicates decreasing economies of scale and $\lambda_c = 1$, constant economies of scale. Increasing economies of scale imply that optimal size exceeds actual size; diseconomies of scale imply that the optimal scale lies below the actual one.

3. Data and Variables

A sample of all 26 Swiss cantons covering the years 2000 to 2004 serves to calculate scale efficiency scores for four public services. These are (1) administration, (2) education, (3) health, and (4) transportation as classified by the Swiss Federal Statistical Office. These four public services are controlled by the cantons. In order to exclude spillovers as far as possible, only primary and secondary education (without tertiary and vocational education, which are in the domain of federal government) and private road transportation (without regional public transportation, for the same reason) are included in the analysis. Unfortunately, adjustments for spillovers that are known to exist especially in health care were not possible.

3.1. Output Variables

Aggregation of output variables proceeds as in Widmer and Zweifel (2012). The four service categories contain between two and six activity-based indicators (see Table 1). Administration is related to population size and the number of registered firms. Education includes four measures for the number of students according to school level. Health is based on case-mix adjusted

Table 1: Output Indicators for the Four Public Services Investigated

Public Service	Output	Description, remarks
(1) Administration		
Legislative, executive General administration	Population No. firms	Population served and number of firms serve as proxies for administration services provided.
(2) Education		
Kindergarden	No. students	The numbers of kindergarden, primary, secondary, and high school students serve as indicators of output values. Quality of education is neglected because of lacking data.
Primary education	No. students	
Secondary education	No. students	
High school education	No. students	
(3) Health		
Hospitals (specialized)	No. patient cases	Case-mix adjusted number of cases serve as a severity-adjusted output for specialized and primary hospitals. The output of rehabilitation and psychiatric clinics, nursing homes and retirement homes is measured by the number of patient days.
Hospitals (primary)	No. patient cases	
Hospitals (psychiatric)	No. patient days	
Rehab clinic	No. patient days	
Nursing homes	No. patient days	
Retirement homes	No. patient days	
(4) Transportation		
Cantonal roads	Road length (km)	Road length serves as a proxy for maintenance.
Communal roads	Road length (km)	Number of cars is used as an indicator of
Road utilization	No. cars	intensity of road use.

patient cases for primary and specialized hospitals and patient days for psychiatric hospitals, rehabilitation clinics, nursing homes, and retirement homes; unfortunately, no adjustment for quality is possible due to lack of data. Finally, transportation incorporates the road length of cantonal and communal roads as well as the number of cars (as an indicator of intensity of use).

3.2. Input Variables

Inputs are aggregated into real expenditure (Swiss francs of 2000) on the four components. This is widespread practice [see Afonso et al. (2006) and De Borger and Kerstens (1996)]. For the components transportation and health, only operating expenses are included (defined as total expenditure minus investment in new infrastructure). Inclusion of investment outlay would have injected a large transitory component into these two categories since cantonal investment cycles differ heavily.

3.3. Data Limitations

Clearly the choice of output variables has an important influence on the results of a DEA. Unfortunately, in public good provision, this choice is no simple task for at least two reasons. First, most of the outputs are not directly quantifiable. Second, public services are too many to

be individually included in a DEA. In the case of Switzerland, most public service outputs are either not observable or not directly quantifiable. In order to nevertheless perform efficiency measurements, Widmer and Zweifel (2012) follow a multi-step procedure. First, for each of the four services they determine a set of activity-based indicators, e.g. number of students for education. Again, this is common procedure [see e.g. Afonso et al. (2006)]. Second, because DEA can accommodate only a limited number of outputs, which in addition should be positively related to inputs, output indicators are derived from a cost driver analysis using seemingly unrelated regression estimates (SURE) for the four services investigated. For these regressions, a panel data set comprising five years was available. The goal was to identify those variables which have the highest positive impact on public expenditure while maximizing the explanatory power of the corresponding model. These criteria permitted to reduce the number of activity-based indicators to two to six variables, depending on the type of service. Finally, Widmer and Zweifel (2012) use DEA weights derived from service-specific DEA models to aggregate the two to six variables to one of the four performance indicators listed in Table 1. In this way, the total number of variable could be further reduced to model the 26 cantonal observations.

The output variables are those that were identified as relevant in the cost driver analysis (see Table 2 again). Although this set is incomplete and cannot fully explain (expenditure on) public services, these variables do explain more than 95 percent of variation in expenditure [see Widmer and Zweifel (2012) once more]. This is sufficient for applying DEA to the provision of public goods. However, results relating to an individual canton should be treated with caution, a major reason being the unobservable volatility in infrastructure investment over time. In an attempt to increase robustness, average values calculated over the years 2000 to 2004 enter the DEA, in contradistinction to Widmer and Zweifel (2012), who use annual data covering 2000 to 2004, too.

4. Results of the DEA

This section first discusses so-called bias-corrected DEA efficiency scores for each of the 26 Swiss cantons and the four types of public service. Each of the four services is analyzed assuming CRS or VRS as explained in Section 2. Robust efficiency scores (see Section 2 again) are derived from an input-oriented DEA. Second, this section contains tests for economies of scale.

This calls for the determination of confidence intervals for both CRS and VRS efficiency scores from consistently estimated density functions. The asymptotic distribution of DEA scores was first constructed by Gijbels et al. (1999) for the case of one input and one output. The homogeneous bootstrapping method proposed by Simar and Wilson (1998) for multiple inputs and outputs is applied here, with bandwidth of 0.19 and 0.17 for the VRS and CRS model, respectively. Confidence intervals are constructed using the Hall percentile intervals described in Simar and Wilson (2000).

4.1. Bias-corrected DEA Efficiency Scores

Table 2 shows average DEA scores for the years 2000 to 2004. At this point, constant economies to scale are assumed (to be changed below). Therefore, the DEA indicates potential cost improvements achievable by a radial movement to a technically and scale-efficient reference point on the frontier. Two super-efficient cases were found in the provision of health; they are assigned a score of 1.00 (see Section 3 again).

Starting with administration, the highest technical efficiency score of 0.98 is attained by the canton of Schwyz (SZ), pointing to an expected potential for cost savings of two percent and between one and seven percent with probability of 95 percent. Among all 26 cantons, there are three more with an expected efficiency score higher than 0.90, viz. Nidwalden (NW, 0.93), Appenzell Inner-Rhodes (AI, 0.92), and St. Gallen (SG, 0.90). It is interesting to note that these four cantons are all rural and small in terms of population and output volume, while cantons below the average value of 0.74 typically are urban and large. The most inefficient canton is Geneva (GE, 0.36); it could lower its expenditure on administration by about 64 percent without reducing output volume.

Turning to education, the CRS efficiency scores show a different picture. Cantons which are efficient in administration need not be efficient producers of educational services. Here, Ticino (TI) ranks highest with a score of 0.99, with the 95 percent confidence interval indicating a cost saving potential between one and four percent. Two more cantons Argovia (AG, 0.98) and Vaud (VD, 0.98) come close. Across the 26 cantons, 12 fail to reach the (high) average value of 0.86, among them also rural and small ones (UR, 0.82; SZ, 0.82; NW, 0.80). The minimum score of 0.66 is associated with the urban canton of Zurich (ZH), the largest canton in terms of population and the economic powerhouse of the country.

Table 2: Bias Corrected Efficiency Scores and Confidence Intervals, 2000 to 2004

Cantons	Administration		Education		Health ¹		Transportation	
	EFF	2.5/ 97.5%	EFF	2.5/ 97.5%	EFF	2.5/ 97.5%	EFF	2.5/ 97.5%
ZH	0.74	(0.71, 0.76)	0.66	(0.65, 0.67)	0.71	(0.68, 0.75)	0.44	(0.45, 0.50)
BE	0.86	(0.83, 0.89)	0.76	(0.74, 0.77)	0.76	(0.72, 0.80)	0.68	(0.65, 0.72)
LU	0.63	(0.62, 0.66)	0.87	(0.85, 0.88)	0.61	(0.60, 0.66)	0.61	(0.59, 0.65)
UR	0.70	(0.68, 0.73)	0.82	(0.80, 0.83)	0.69	(0.66, 0.73)	0.64	(0.62, 0.69)
SZ	0.98	(0.93, 0.99)	0.82	(0.81, 0.83)	0.60	(0.58, 0.64)	0.67	(0.64, 0.71)
OW	0.78	(0.75, 0.80)	0.94	(0.92, 0.95)	0.66	(0.63, 0.70)	0.68	(0.65, 0.72)
NW	0.93	(0.88, 0.95)	0.80	(0.78, 0.81)	0.59	(0.58, 0.64)	0.70	(0.68, 0.75)
GL	0.73	(0.71, 0.76)	0.84	(0.83, 0.86)	0.94	(0.87, 0.97)	0.61	(0.59, 0.65)
ZG	0.87	(0.83, 0.89)	0.66	(0.65, 0.67)	0.80	(0.76, 0.84)	0.56	(0.55, 0.61)
FR	0.79	(0.76, 0.81)	0.92	(0.90, 0.93)	0.68	(0.65, 0.72)	0.65	(0.63, 0.69)
SO	0.81	(0.78, 0.84)	0.85	(0.83, 0.86)	0.63	(0.61, 0.67)	0.82	(0.78, 0.86)
BS	0.54	(0.54, 0.58)	0.80	(0.78, 0.81)	0.60	(0.58, 0.64)	0.32	(0.36, 0.39)
BL	0.65	(0.63, 0.68)	0.82	(0.80, 0.83)	0.68	(0.65, 0.72)	0.53	(0.53, 0.58)
SH	0.69	(0.67, 0.72)	0.82	(0.80, 0.83)	0.67	(0.65, 0.72)	0.93	(0.87, 0.96)
AR	0.69	(0.67, 0.72)	0.90	(0.88, 0.91)	0.96	(0.90, 0.99)	0.55	(0.54, 0.60)
AI	0.92	(0.87, 0.93)	0.91	(0.89, 0.92)	1.00	(1.00, 1.00)	0.59	(0.57, 0.63)
SG	0.90	(0.86, 0.92)	0.82	(0.80, 0.83)	0.76	(0.72, 0.80)	0.42	(0.43, 0.48)
GR	0.60	(0.59, 0.63)	0.87	(0.85, 0.88)	0.67	(0.65, 0.71)	0.47	(0.48, 0.53)
AG	0.78	(0.75, 0.80)	0.98	(0.95, 0.99)	0.66	(0.64, 0.70)	0.91	(0.85, 0.94)
TG	0.85	(0.81, 0.87)	0.88	(0.86, 0.89)	1.00	(1.00, 1.00)	0.97	(0.90, 0.99)
TI	0.60	(0.59, 0.63)	0.99	(0.96, 0.99)	0.77	(0.73, 0.80)	0.82	(0.77, 0.85)
VD	0.64	(0.63, 0.67)	0.98	(0.96, 0.99)	0.60	(0.58, 0.64)	0.63	(0.61, 0.68)
VS	0.85	(0.81, 0.87)	0.95	(0.93, 0.96)	0.85	(0.80, 0.88)	0.79	(0.75, 0.83)
NE	0.66	(0.64, 0.68)	0.96	(0.93, 0.97)	0.36	(0.39, 0.44)	0.45	(0.46, 0.51)
GE	0.36	(0.38, 0.41)	0.87	(0.85, 0.88)	0.90	(0.85, 0.94)	0.62	(0.60, 0.67)
JU	0.72	(0.70, 0.75)	0.88	(0.86, 0.89)	0.37	(0.40, 0.44)	0.90	(0.85, 0.93)
Mean	0.74		0.86		0.69		0.66	
Min	0.37		0.66		0.37		0.32	
Max	0.98		0.99		1.00		0.97	

¹ AI and TG are identified as outliers and are excluded from the production technology set Γ .

Health comes up with different efficiency scores again. Appenzell Inner-Rhodes (AI, 1.00) and Thurgovia (TG, 1.00) are the two cantons that are identified as super-efficient. This implies that they do not serve as comparators for the others. With this qualification, rural and small Appenzell Outer-Rhodes (AO, 0.96) comes in first, followed by Glarus (GL, 0.94). Calculated 95 percent confidence intervals indicate cost saving potentials between one to ten percent and three to 13 percent, respectively. Mean CRS efficiency is 0.69, somewhat lower than in administration (0.74) and much lower than in education (0.86). Also, no less than 14 cantons fail to achieve this average, indicating negative skewness. Indeed, at the lower end, Neuchâtel (NE, 0.36) and Jura (JU, 0.37) are very far below average.

Finally, in transportation, Thurgovia (TG, 0.97) again ranks first (as in health). With scores of 0.85 in administration and 0.88 in education, this rural and medium-sized canton has highest overall efficiency. Three more cantons have transportation scores higher than 0.90, viz. Schaffhausen (SH, 0.93), Argovia (AG, 0.91) and Jura (JU, 0.90). All the other cantons have a mean efficiency score of 0.66, resulting in cost saving potentials of ten percent and higher. Across the four types of service, transportation has the lowest average value, implying that the Swiss cantons could have saved about 34 percent of their public expenditure without jeopardizing the provision of transportation services.

The scores of Table 2 prove to be quite similar to those presented by Widmer and Zweifel (2012) based on annual rather than averaged data. In particular, Thurgovia is most efficient while Basel-City is least efficient in both studies. In addition, the four types of service present a similar picture in that cantons seem to be relatively efficient and homogenous in their provision of educational services. Thus, the claim that lack of coordination between them causes inefficiency continues to lack empirical support. Indeed, education turns out to perform much better than health and transportation and also better than administration, a service in which Switzerland is generally believed to excel. Thus, even though the present analysis adjusts for volatility in infrastructure and corrects for possible estimation bias, the policy conclusions remain the same as in Widmer and Zweifel (2012).

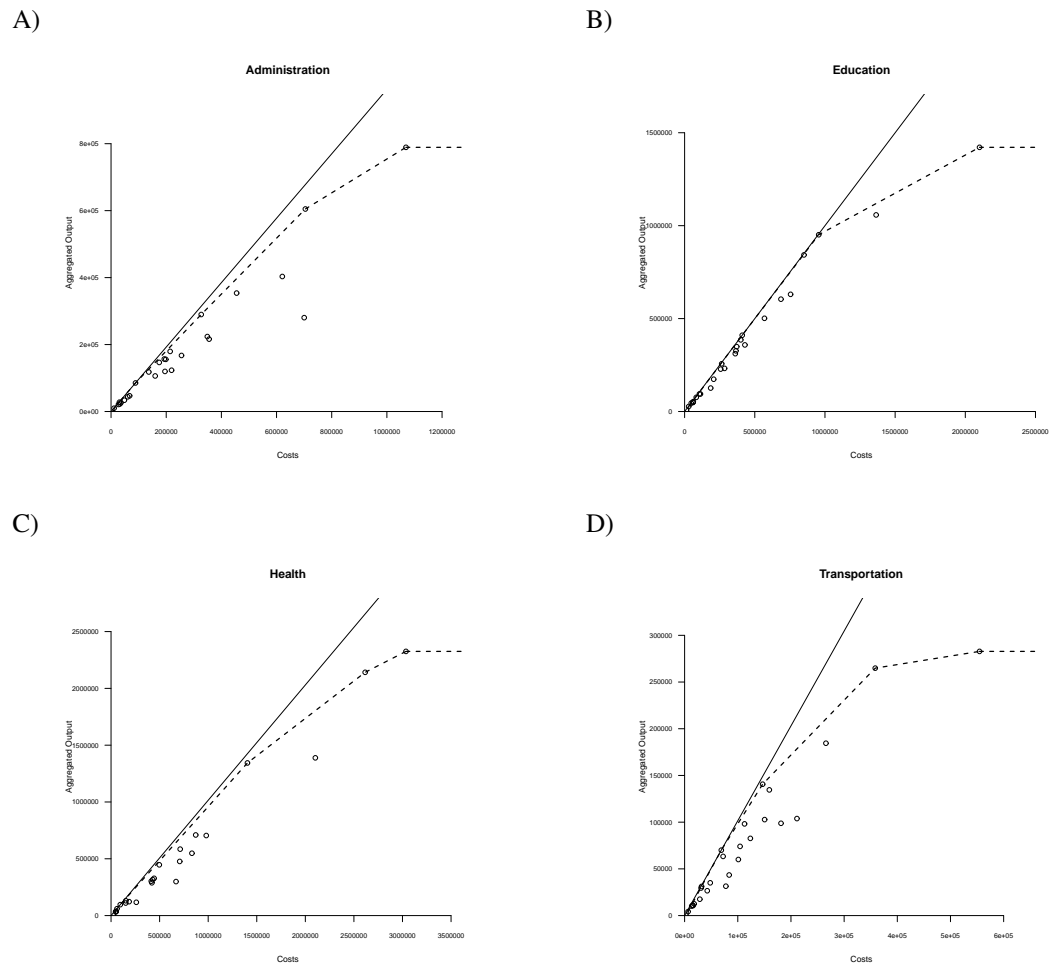
4.2. Optimal Scale of Public Good Provision

This subsection addresses the question of whether the cost saving potentials of Table 1 reflect scale inefficiencies. Economies of scale are said to exist if a canton can provide a given

level of service at a lower public expenditure than other cantons of the same size. Diseconomies of scale exist if a canton can provide a given level of service only at a higher expenditure than others (see Section 3 again). To test for these effects, the bootstrapped CRS efficiency scores of Table 1 are compared to their VRS counterparts [see equation (3) again]. If a canton is found to perform better under the VRS assumption than under the CRS assumption, it has potential to improve its performance by adjusting its scale of provision [see equation (4) again].

Local economies and diseconomies of scale are apparent in Figure 1 because the CRS efficient frontier (solid line) diverges from its counterpart (dashed line) for the four services considered (for the shares of cantons exhibiting increasing, constant, and decreasing economies of scale, see Table 3).

Figure 1: Bias-corrected scale efficiencies



Solid line: constant economies of scale (CRS), dashed line: variable economies of scale (VRS)

Starting once more with administration, panel A of Figure 1 shows that the VRS and CRS frontiers coincide for relatively small cantons, calling for modest reductions in size to achieve full efficiency. As evidenced in Table 3, no canton exhibits increasing economies of scale. The majority of cantons operates at constant (57.7 percent) or decreasing (42.3 percent) economies of scale. Given that even very small cantons are among the 57.7 percent with CRS, there are no apparent efficiency reasons for their consolidation into larger units.

In the case of education, panel B of Figure 1 presents a somewhat different picture. The VRS and CRS efficiency frontiers again coincide, but this time over almost the entire range of cantons, pointing to an overall absence of economies and diseconomies of scale in the provision of educational services. Table 3 confirms these findings in that 65.4 percent of all cantons have optimal size (constant economies of scale) while only 11.5 percent could benefit from increasing economies of scale. Conversely, 23.1 percent operate at an excessive scale. Therefore, the very small cantons might achieve efficiency gains by cooperating in primary and secondary education, while the largest cantons might consider increased delegation of authority to their districts and communes (see below).

Turning to health, panel C of Figure 1 shows a situation similar to administration. The CRS and VRS efficiency frontiers can hardly be distinguished and remain close together for cantons of medium and large size. However, Table 3 does indicate that only 16.7 percent of cantons exhibit increasing economies of scale, while 45.8 percent have decreasing economies of scale. This result shows that health care services should be more decentralized, contrary to the mantra of current public policy arguing that a more centralized provision of e.g. hospital services would permit efficiency gains. The findings of this study are in line with those of Farsi and Filippini (2006), who were unable to identify large economies of scale in the Swiss hospital sector.

Transportation is the only service that seems to be characterized by increasing returns to scale, with panel D of Figure 1 indicating that the VRS efficiency frontier diverges from the CRS one for small cantons already. Medium-sized cantons operate at increasing returns to scale, likely because this type of service is extremely capital intensive. According to Table 3, 38.5 percent of all cantons might benefit from increasing scale. However, 53.8 percent of all cantons operate with constant economies of scale and thus have optimal size.

A comparison across the four types of public service reveals differences in the structure of production. While administration calls for a very decentralized provision for scale efficiency,

Table 3: Share of Cantons with Increasing, Constant, and Decreasing Economies of Scale (with Bias Correction)

Public service	below optimal scale	at optimal scale ¹	above optimal scale
Administration	0.0	57.7	42.3
Education	11.5	65.4	23.1
Health	16.7	37.5	45.8
Transportation	38.5	53.8	7.7

¹ Defined as scale efficiency score of 0.99 and higher.

health and transportation are optimally provided by cantons of a medium size.

Referring again to the decentralization principle of Oates (1972) and FOCJ proposed by Frey (2005), these suggestions are supported by the findings at least in the case of administration and education. A very decentralized provision may well be welfare maximizing, provided it matches consumer preferences as closely as possible. However, even for the other two services, an assignment of authorities and responsibilities exclusively based on the argument that economies of scale exist is not justified. The benefits to be reaped by operating at a larger scale in transportation and possibly health (see Table 3 again) still would have to be weighted against the welfare losses caused by a lesser degree of adaptation to citizen preferences and reduced citizen control of government that usually go along with an increased size of jurisdiction Oates (2006).

The FOCJ concept appears to be useful in this context. Especially in the domain of transportation, small cantons might create a joint institution for planning, construction, and maintenance while retaining their autonomy in the provision of administration and educational services. In the case of health care, smaller cantons would be well advised to opt for more co-operation; by way of contrast, the large ones would benefit from even more decentralization by delegating authority to districts and communes. Still, consumer preferences have to be taken into account to find the optimal level of provision. Failure to do this would cause citizens to move away from their canton of residence in search of another offering a more favorable benefit-cost ratio in the provision of public services (Tiebout competition). At least in the case of Switzerland, Tiebout competition does not entail much cost because shrinking jurisdictions would not forgo increasing returns to scale (with some exceptions in the case of transportation) while growing jurisdictions are unlikely to suffer from decreasing returns (with exceptions in administration and health).

These statements are derived from findings that need to be checked for their robustness. One

check is to calculate efficiency scores that are not bias-corrected. Table 4 presents the comparison. While according to the bias-corrected DEA scores, no canton has increasing economies of scale in administration, the uncorrected scores indicate a share of 23.1 percent in this category. And rather than 57.7 percent, now only 11.5 percent of them exhibit constant returns to scale. In the case of transportation, the shift away from the CRS category is of the same magnitude (from 53.8 to 15.4 percent); it also occurs more mildly in the case of education and health. Put the other way around, failure to correct DEA efficiency scores for bias results in an overestimation of the benefits to be achieved from changing the scale at which cantons are operating. This increased stability is the consequence of the fact that with bias correction, the location of the efficient frontier is less likely to depend on a single observation, causing it to become smother and more robust against outliers that suggest economies of scale.

Table 4: Share of Cantons with Increasing, Constant, or Decreasing Economies of Scale (without Bias Correction²)

Public service	below optimal scale		at optimal scale ¹		above optimal scale	
	Bias corr.	No bias corr.	Bias corr.	No bias corr.	Bias corr.	No bias corr.
Administration	0.0	23.1	57.7	11.5	42.3	65.4
Education	11.5	34.6	65.4	57.7	23.1	7.7
Health	16.7	16.7	37.5	8.3	45.8	75.0
Transportation	38.5	46.2	53.8	15.4	7.7	38.5

¹ Defined as scale efficiency score of 0.99 and higher.

² For bias-corrected values see Table 3.

4.3. Testing for the Existence of Global Economies of Scale in the Provision of Public Goods

Up to this point, the analysis has focused on the local existence of returns to scale. This leaves open the possibility of scale effects across all Swiss cantons, small, medium-sized, and large. If economies of scale do exist, the hypothesis of constant economies of scale (CRS) would have to be rejected in favor of the VRS alternative. The null hypothesis H_0 therefore is that the technology set Γ (see Section 2) is CRS, while H_A is defined as Γ being VRS. For testing, the approach proposed by Bogetoft and Otto (2011) is employed by calculating the test statistic

$$S = \frac{\sum_{c=1}^C E_{CRS}^c}{\sum_{c=1}^C E_{VRS}^c}, \quad (5)$$

where E_{CRS}^c symbolizes the non-bootstrapped efficiency score of canton c under the CRS hypothesis. This statistic is compared with a critical value of the distribution of efficiency scores obtained from bootstrapping. If S is below the critical value, then H_0 is rejected, because the less restrictive VRS scores exceed the CRS ones by "too little". Table 5 displays the values of the test statistic S along with the corresponding critical values at a significance level α of 0.05.

Table 5: Test Statistics of H_0 : Technology is CRS

Public service	test statistic S	critical value ($\alpha = 0.05$)
Administration	0.94	0.89
Education	0.97	0.94
Health	0.91	0.86
Transportation	0.89	0.85

As Table 5 shows, S exceeds the critical value for all four public services, suggesting acceptance of H_0 . Thus, constant economies of scale prevail, confirming the analysis of local effects to scale in Section 4.2. This means that centralized provision of public services or consolidation of local jurisdictions cannot be justified in general; rather, for each canton and type of service, optimal scale needs to be individually determined.

5. Conclusions

The assignment of responsibilities to levels of government is an ongoing issue among federal states worldwide. The question is, "What allocation of authorities and responsibilities is most likely to maximize national welfare?". Although the economic literature on fiscal federalism provides reasoned answers to this question, consolidation of jurisdictions is nowadays driven by the expectation of reaping scale efficiencies only.

This article tests whether this expectation holds true in the case of the 26 cantons of Switzerland. Using averaged annual data covering the years 2000 to 2004, it applies Data Envelopment Analysis (DEA) to four types of public service (administration, education, health, and transportation). Bias-corrected DEA scores for technical and scale efficiency are derived for each canton to find out whether these four services are characterized by nonconstant economies of scale.

The main results are the following. First, significant cost efficiency differences exist among the 26 Swiss cantons and between the four public services analyzed. A canton that is efficient in the provision of one public service need not be efficient in the three other activities. In general, small rural cantons perform better than large urban ones. Second, not one of the four public services exhibits overall economies of scale. This finding lends support to voters' resistance against mergers between cantons in the name of increased efficiency (they rejected the proposed merger between the cantons of Geneva and Vaud in 2002). Third, health and transportation are optimally provided by cantons of medium size, while administration and education call for a very decentralized provision for scale efficiency. Thus, the 'natural monopoly' argument does not apply, leaving scope for Tiebout competition between the cantons.

This analysis suffers from several limitations. Above all, DEA efficiency scores constitute a very technocratic measure, being silent on the question of whether the services provided reflect the preferences of citizens. Also, some of the input and output variables used to calculate efficiency scores may not be fully adequate for modeling the production process of public goods. In particular, health only incorporates hospital services but completely neglects ambulatory care. However, in spite of these limitations, there is something to be learned from this work. As argued in the Introduction section, an efficient structuring of jurisdictions in a federal state calls for a delicate balance between economies of scale on the one hand and citizens' preferences on the other. This balance becomes even more delicate when economies of scale vary between public services, as shown in this contribution. In this situation, the concept of functional, overlapping, and competing jurisdictions (FOCJ) appears to be useful. Still, consumer preferences need to be taken into account to find the optimal level of provision. Failure to do this would cause citizens to move away from their canton of residence in search of another jurisdiction offering a more favorable benefit-cost ratio in the provision of public services. However, this Tiebout competition does not entail much opportunity cost because most shrinking jurisdictions would not forgo returns to scale while growing ones would continue to operate under constant returns to scale.

6. References

References

- Afonso, A., Schuknecht, L., Tanzi, V., 2006. Public sector efficiency: Evidence for new EU member states and emerging markets. Working Papers 01, Department of Economics at the School of Economics and Management (ISEG), Technical University of Lisbon.
- Andersen, P., Petersen, N. C., 1993. A procedure for ranking efficient units in data envelopment analysis. *Management Science* 39 (10), 1261–1264.
- Anwar, S., 2004. Fiscal decentralization in developing and transition economies: Progress, problems, and the promise. World Bank Policy Research Working Paper 3282, World Bank.
- Baleiras, R. N., 2001. To fragment or to consolidate jurisdictions: the optimal architecture of government. FEUNL Working Paper Series wp401, Universidade Nova de Lisboa, Faculdade de Economia.
- Banker, R., Charnes, A., Cooper, W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30 (9), 1078–1092.
- Bogetoft, P., Otto, L., 2011. *Benchmarking with DEA, SFA, and R*. Springer.
- Charnes, A., Cooper, W., Rhodes, E., 1979. Measuring the efficiency of decision-making units. *European Journal of Operational Research* 3 (4), 339–338.
- De Borger, B., Kerstens, K., 1996. Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics* 26 (2), 145–170.
- Farrell, M., 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society* 120 (3), 253–281.
- Farsi, M., Filippini, M., 2006. An analysis of efficiency and productivity in Swiss hospitals. *Swiss Journal of Economics and Statistics (SJES)* 142 (I), 1–37.
- Frey, B. S., 2005. Functional, overlapping, competing jurisdictions: Redrawing the geographic borders of administration. *European Journal of Law Reform* 3 (4), 543–555.

- Gijbels, I., Mammen, E., Park, B. U., Simar, L., 1999. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association* 94 (445), 220–228.
- Gómez, J. L., 2010. Essays on optimal jurisdictional size for local service delivery. Dissertation, Georgia State University, Institute of Technology.
- Kellermann, K., 2007. Die öffentlichen Ausgaben der Kantone und ihrer Gemeinden im Quervergleich (Cross-sectional comparison of public expenditures between Swiss cantons). Working Paper 37, Federal Department of Economic Affairs DEA, Switzerland.
- Koopmans, T., 1951. An analysis of production as an efficient combination of activities. In: Koopmans, T. (Ed.), *Activity Analysis of Production and Allocation*. John-Wiley and Sons, Inc., New York, pp. 33–97.
- Oates, W., 1972. *Fiscal Federalism*. NY:Harcourt Brace Jovanovich, Princeton.
- Oates, W., 2006. On the theory and practice of fiscal decentralization. Working Papers 05, Institute for Federalism & Intergovernmental Relations.
- Olson, M. J., 1969. The principle of fiscal equivalence: The division of responsibilities among different levels of government. *The American Economic Review* 59 (2), 479–487.
- Simar, L., Wilson, P. W., 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science* 44, 49–61.
- Simar, L., Wilson, P. W., 2000. Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis* 13, 49–78.
- Widmer, P., Zweifel, P., 2012. Fiscal equalization, tiebout competition, and incentives for efficiency in a federalist country. *Public Finance Review* 40 (1), 3–29.